

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: December 9, 2000, 10:18:13 ; Search time 4054.2 seconds
(without alignments)
3971.996 Million cell updates/sec

Title: US-09-117-447-1
Perfect score: 3687
Sequence: 1 atgagataagaaagaaagctgt.....ttacattctatataagtaa 3687

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_cm:*
4: gb_cv:*
5: gb_pac:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pl4:*
11: gb_pl5:*
12: gb_pl6:*
13: gb_pl7:*
14: gb_pl8:*
15: gb_pl9:*
16: gb_pl10:*
17: gb_pl11:*
18: gb_pl12:*
19: gb_pl13:*
20: gb_pl14:*
21: gb_pl15:*
22: gb_pl16:*
23: gb_pl17:*
24: gb_pl18:*
25: gb_pl19:*
26: gb_pl20:*
27: gb_pl21:*
28: gb_pl22:*
29: gb_pl23:*
30: gb_pl24:*
31: gb_pl25:*
32: gb_pl26:*
33: gb_pl27:*
34: gb_pl28:*
35: gb_pl29:*
36: gb_pl30:*
37: gb_pl31:*
38: gb_pl32:*
39: gb_pl33:*
40: gb_pl34:*
41: gb_pl35:*
42: gb_pl36:*
43: gb_pl37:*

44: em_hcg1:*
45: em_hcg2:*
46: em_hcg3:*
47: em_hcg4:*
48: em_hcg5:*
49: em_hcg6:*
50: em_hcg7:*
51: em_hcg8:*
52: em_hcg9:*
53: em_hcg10:*
54: em_hcg11:*
55: em_hcg12:*
56: em_hcg13:*
57: em_hcg14:*
58: em_hcg15:*
59: em_hcg16:*
60: em_hcg17:*
61: em_hcg18:*
62: em_hcg19:*
63: em_hcg20:*
64: em_hcg21:*
65: em_hcg22:*
66: em_hcg23:*
67: em_hcg24:*
68: em_hcg25:*
69: em_hcg26:*
70: em_hcg27:*
71: em_hcg28:*
72: em_hcg29:*
73: em_hcg30:*
74: em_hcg31:*
75: em_hcg32:*
76: em_hcg33:*
77: em_hcg34:*
78: em_hcg35:*
79: em_hcg36:*
80: em_hcg37:*
81: em_hcg38:*
82: em_hcg39:*
83: em_hcg40:*
84: em_hcg41:*
85: em_hcg42:*
86: em_hcg43:*
87: em_hcg44:*
88: em_hcg45:*
89: em_hcg46:*
90: em_hcg47:*
91: em_hcg48:*
92: em_hcg49:*
93: em_hcg50:*
94: em_hcg51:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3687	100.0	3687	1 AX000218	AX000218 Sequence 1
2	3687	100.0	3687	2 A93832	A93832 Sequence 1
3	3687	100.0	3687	2 BSS858	X71092 B. stea:ctc
4	513	13.9	3680	1 AF055578	AF055578 Bacillus
5	487.6	13.2	3090	1 AF228338	AF228338 Bacillus
6	77.4	2.1	2069	22 E10125	E10125 DNA encodin
7	77.4	2.1	3399	22 E10126	E10126 DNA encodin
8	77	2.1	11554	8 AP002460	AP002460 Arabidops
9	73	2.0	5420	33 AF056936	AF056936 Plasmodi
10	71.4	1.9	17953	52 AC024253	AC024253 Homo sapi
11	64.2	1.7	110000	71 PFMALP1_1	Continuation (2 of
12	64	1.7	3720	73 S76368	S76368 ORF 5' of E

C	13	64	1.7	43658	73	HSV3P3RCGN
C	14	64	1.7	112930	73	HSGEND
C	15	62.6	1.7	12029	32	AE0013373
C	16	62.4	1.7	1086	22	EO8995
C	17	61.6	1.7	423	2	MC049
C	18	60.2	1.6	4050	57	DDPFRKING
C	19	59	1.6	1324	57	PERESARI
C	20	59	1.6	1587	57	PFAL18528
C	21	59	1.6	4591	5	AO0661
C	22	58.8	1.6	4995	57	PR4MSA
C	23	57.8	1.6	2196	2	BBU78764
C	24	57.8	1.6	2261	2	BBU72896
C	25	57.8	1.6	2642	33	AEI03869
C	26	57.8	1.6	30750	1	AE001575
C	27	57.8	1.6	30885	1	AE001580
C	28	57.6	1.6	186935	51	AC0223322
C	29	57.4	1.6	4283	2	BBU44912
C	30	56	1.5	189717	54	AC026679
C	31	55.6	1.5	152409	71	PFMAL1P1
C	32	55.4	1.5	169676	52	AC023334
C	33	55	1.5	5173	57	DDG14DS
C	34	55	1.5	6797	2	MHMP34GN
C	35	55	1.5	13674	57	U00691
C	36	54.8	1.5	151922	30	AC0062278
C	37	54.2	1.5	9200	57	MITBOX
C	38	54.2	1.5	23016	57	TBRKXGN
C	39	54	1.5	171627	50	AC020602
C	40	53.4	1.4	7218	5	I66494
C	41	53	1.4	4102	33	AE270648
C	42	53	1.4	11650	74	U32846
C	43	52.6	1.4	397	93	G37798
C	44	52.2	1.4	2384	57	PFAGNR
C	45	52.2	1.4	1839	48	SCU12825
C	13	64	1.7	43658	73	HSV3P3RCGN
C	14	64	1.7	112930	73	HSGEND
C	15	62.6	1.7	12029	32	AE0013373
C	16	62.4	1.7	1086	22	EO8995
C	17	61.6	1.7	423	2	MC049
C	18	60.2	1.6	4050	57	DDPFRKING
C	19	59	1.6	1324	57	PERESARI
C	20	59	1.6	1587	57	PFAL18528
C	21	59	1.6	4591	5	AO0661
C	22	58.8	1.6	4995	57	PR4MSA
C	23	57.8	1.6	2196	2	BBU78764
C	24	57.8	1.6	2261	2	BBU72896
C	25	57.8	1.6	2642	33	AEI03869
C	26	57.8	1.6	30750	1	AE001575
C	27	57.8	1.6	30885	1	AE001580
C	28	57.6	1.6	186935	51	AC0223322
C	29	57.4	1.6	4283	2	BBU44912
C	30	56	1.5	189717	54	AC026679
C	31	55.6	1.5	152409	71	PFMAL1P1
C	32	55.4	1.5	169676	52	AC023334
C	33	55	1.5	5173	57	DDG14DS
C	34	55	1.5	6797	2	MHMP34GN
C	35	55	1.5	13674	57	U00691
C	36	54.8	1.5	151922	30	AC0062278
C	37	54.2	1.5	9200	57	MITBOX
C	38	54.2	1.5	23016	57	TBRKXGN
C	39	54	1.5	171627	50	AC020602
C	40	53.4	1.4	7218	5	I66494
C	41	53	1.4	4102	33	AE270648
C	42	53	1.4	11650	74	U32846
C	43	52.6	1.4	397	93	G37798
C	44	52.2	1.4	2384	57	PFAGNR
C	45	52.2	1.4	1839	48	SCU12825
C	13	64	1.7	43658	73	HSV3P3RCGN
C	14	64	1.7	112930	73	H

ALIGNMENTS

[illegible]

```

RESULT 1
AX000218
LOCUS AX000218 3687 bp DNA
DEFINITION Sequence 1 from Patent WO9906567.
ACCESSION AX000218
VERSION AX000218.1 GI:7240682
KEYWORDS
SOURCE
ORGANISM
Bacillus stearothermophilus.
Bacillus stearothermophilus
Bacteria; Firmicutes; Bacillus/lostiridium group;
Bacillus/streptococcus group; Bacillus.
1 (bases 1 to 3687)
Lubitz W. and Resch S.
EXTRACTED FROM PATENT WO 9906567-A 11-FEB-1999;
LUBITZ WERNER (AT); RESCH STEPHANIE (DE)
location/Qualifiers
1.3687
/organism="Bacillus stearothermophilus"
/strain="PV72"
/DB_xref="taxon:1422"
/clone="SBSA"
1.90
1.3687
sig_peptide-
CDS

```

```

/note="unnamed protein product"
/codon_start=1
/transl_table=11
/translation_id="CAB7069.1"
/translation="MDRRKAVKLATASATAASAFAVPAANPNASEATDAIVAYSGAKQK
EKAYATAYTSTVETGEPPNIDVYAEYNNKRTPRVALYNNAGSKRATYADLC
KEETVYKAMPNGSGEABVANNYIDAYVATVLDIMBRELAAVOAKDLERKQYVYHK
PYELKTYVILDRYGGKTRDLNLSRTEKRAQELRDSLLIYITAMAKREYQDAVYAK
NDRKAAYVOIINOYLERVYDAETELTEVAKKILMDEALTEKVESVSAKLINONKKA

```

	mat_peptide	BASE COUNT	1316 a	660 c	715 g	996 t	
	ORIGIN						
	Query Match	100.0%;	Score 3687;	DB 1;	Length 3687;		
	Best Local Similarity	100.0%;	Pred. No. 0;				
	Matches 3687;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1 atggatgagaaaaaagcctgtgaactgacacacgaagctgtatctgcgcagcagtcatt	60					
DB	1 ATGGATAGGAAAAAGCGCTGTGAACATACCAACGCAAGTGGTATTCGAGCAAGTCATT	60					
QY	61 gtgcctgcacatcccaacgctctgaagcgcgtacagatgtgacacagatgaagccaa	120					
DB	61 GTCGCTGCMAATCCAAAGCGCTTCTGAACCGGCTACAGATGTGACCAACGTTGAAGCCAA	120					
QY	121 gcaaaagcacagtlccaaaaagcatcattactcttaacagccatlacagtaacggaac	180					
DB	121 GCAAAAAGCAGATWCAAAAAAGACATATATCTTACGCCATACGTAACGGAACCTGCT	180					
QY	181 gaattcccaaacattaaacgagtatactgcgtgaatcaacaaagcgaaaaaacgatac	240					
DB	181 GAATTCCCAACCATTAAGATGATATATCTGAATTCACAAACGAAGCCAAAAAGCATAC	240					
QY	241 gctgctgagacatctgaataaagcaggctggcgaaaaaagacgcttacttaagctat	300					
DB	241 GATGGCGGATGACATTAATGATTAACCAAGATGGCGGAAAAAGACGCTTACTTACATAT	300					
QY	301 ttacaaaaagatatgaacttcgcttttcbaagcnaacccctaaatcttgcggaagctc	360					
DB	301 TTACAAAAAGATATGAAGATTCAGTATTCGTTTTCANAGCAAAACCCGTAATTCGGCAACTG	360					
QY	361 gtacgaacttaccatgcagtcttaacaactatgcacanaattagacgaatgcgcgaag	420					
DB	361 GTACGAACCTTACATGATGAGCTTTCACACATATGCMAAAAAATTGACGAATTCGCCAAAG	420					
QY	421 ctgaaggcttcgcttcaagcgaagaattagaaaaaaggaagaacatattatcaaaaa	480					
DB	421 CTGAGGCGCTGCTTCAAGCAGAAATTTAGAAAAAGCAACAATTAATTCACAAAAT	480					
QY	481 ccttgaanaatnaactgcgcagctcaattttagatcgcgataatggtaaaaaacactc	540					
DB	481 CCTTATGAATTAATAAAGCTGCACATCTTATTAAGTTCGCGTATATGTTAAAACAATCGT	540					
QY	541 gatttacttcgctctacatttaagcgaagaagcacaagaacttcgcgaagcttaattat	600					
DB	541 GATTTACTTTCGCTCTACATTTAAGCAAAAGCAAGAATTCGCGACAGCTTAATTAT	600					
QY	601 gatattaccgttgcataagaagcgcgcgaagtacaagaagctgttgaagcaggcaact	660					
DB	601 GATATTACCGTTGCATATGAAGCGCGCAAGTACAGAGCGCTGTAAGCAGGCAATTAT	660					
QY	661 gacaaagcnaagcgcgtgttgaatcaaatcaataacttaacaaagaatgaacagctc	720					
DB	661 GACAAAGCTAAAGCGCTGTGATCAATTCATCAATCTTACCAAAAGTAAACAATGCT	720					

OY 721 ttcacaaactgaaactaacagaaagtagcgaacaaagcattagatgcagatgaagctgcgctt 780
DB 721 ttcacaaactgaaactaacagaaagtagcgaacaaagcattagatgcagatgaagctgcgctt 780
OY 781 actcacaagaagtggaagtgtaagtgcgatcaacacccaagaagctgttgattaca 840
DB 781 actcacaagaagtggaagtgtaagtgcgatcaacacccaagaagctgttgattaca 840
OY 841 gaaagcaccgtagaagcgaacactaaattacacattcgcgcgcgcgaatgaatga 900
DB 841 gaaagcaccgtagaagcgaacactaaattacacattcgcgcgcgcgaatgaatga 900
OY 901 gtaaacgtaaatatcgtatgaatctataaagtggaagcgaatccattcgcctta 960
DB 901 gtaaacgtaaatatcgtatgaatctataaagtggaagcgaatccattcgcctta 960
OY 961 accgcaagatgcttcttctatcacaagcgaacaaactacacgtagatgcttcaactca 1020
DB 961 accgcaagatgcttcttctatcacaagcgaacaaactacacgtagatgcttcaactca 1020
OY 1021 ttcgaaataaatacgaagatgaatgaatgaatgaatgaatgaatgaatgaatga 1080
DB 1021 ttcgaaataaatacgaagatgaatgaatgaatgaatgaatgaatgaatgaatga 1080
OY 1081 gaattaaagaagatgcatcacttcaagctcgaatgagatgagatgagatgagatg 1140
DB 1081 gaattaaagaagatgcatcacttcaagctcgaatgagatgagatgagatgagatg 1140
OY 1141 ttggaactaatgtaacaaacacacacacacacacacacacacacacacacacacac 1200
DB 1141 ttggaactaatgtaacaaacacacacacacacacacacacacacacacacacacac 1200
OY 1201 gacgaactaatgaagtagtatgtaaaagtgtgagacacgaacacgaacacgaac 1260
DB 1201 gacgaactaatgaagtagtatgtaaaagtgtgagacacgaacacgaacacgaac 1260
OY 1261 aacgttactatcagaatgctgaacacgaacacgaacacgaacacgaacacgaac 1320
DB 1261 aacgttactatcagaatgctgaacacgaacacgaacacgaacacgaacacgaac 1320
OY 1321 gcttctacaaactaacatcactgctgaacaaagacgctgaacacgaacacgaac 1380
DB 1321 gcttctacaaactaacatcactgctgaacaaagacgctgaacacgaacacgaac 1380
OY 1381 gctacaaataatgtaacacacacacacacacacacacacacacacacacacac 1440
DB 1381 gctacaaataatgtaacacacacacacacacacacacacacacacacacacac 1440
OY 1441 actgcaaacgcatcagcaacacacacacacacacacacacacacacacacacacac 1500
DB 1441 actgcaaacgcatcagcaacacacacacacacacacacacacacacacacacacac 1500
OY 1501 ttaatacctgcttctcctcacaacaaatgcttgaggatgaatgctgagatga 1560
DB 1501 ttaatacctgcttctcctcacaacaaatgcttgaggatgaatgctgagatga 1560
OY 1561 gctggaactaatatcctgcttcaacacgaacacacacacacacacacacacacac 1620
DB 1561 gctggaactaatatcctgcttcaacacgaacacacacacacacacacacacacac 1620
OY 1621 tctactttagctgaataactgttatagtgtaaaagaactgtagcaactgttgcttct 1680
DB 1621 tctactttagctgaataactgttatagtgtaaaagaactgtagcaactgttgcttct 1680
OY 1681 gaactaataatataagcagaacgctaaatgtaactgtagcgaacacgagcctta 1740
DB 1681 gaactaataatataagcagaacgctaaatgtaactgtagcgaacacgagcctta 1740
OY 1741 gaaataataatataatgaacacacacacacacacacacacacacacacacacacac 1800
DB 1741 gaaataataatataatgaacacacacacacacacacacacacacacacacacacac 1800

OY 1801 ttaagcactgttaacgagaacacacacacacacacacacacacacacacacacacac 1860
DB 1801 ttaagcactgttaacgagaacacacacacacacacacacacacacacacacacacac 1860
OY 1861 gttatagagtagtctaaataatgagcgttgatgaatgaatgaatgaatgaatga 1920
DB 1861 gttatagagtagtctaaataatgagcgttgatgaatgaatgaatgaatgaatga 1920
OY 1921 ttaactgtgaatgttcacagaatataacacacacacacacacacacacacacacac 1980
DB 1921 ttaactgtgaatgttcacagaatataacacacacacacacacacacacacacacac 1980
OY 1981 ac 2040
DB 1981 ac 2040
OY 2041 gcaagtgac 2100
DB 2041 gcaagtgac 2100
OY 2101 gtagaagtgctgctac 2160
DB 2101 gtagaagtgctgctac 2160
OY 2161 aagatgac 2220
DB 2161 aagatgac 2220
OY 2221 atctatagcttcaac 2280
DB 2221 atctatagcttcaac 2280
OY 2281 aagatgac 2340
DB 2281 aagatgac 2340
OY 2341 caaaatcactac 2400
DB 2341 caaaatcactac 2400
OY 2401 ggcgaagtaac 2460
DB 2401 ggcgaagtaac 2460
OY 2461 actgatacagctgctcctcagaacacacacacacacacacacacacacacacacac 2520
DB 2461 actgatacagctgctcctcagaacacacacacacacacacacacacacacacacac 2520
OY 2521 ttaactgctcctgac 2580
DB 2521 ttaactgctcctgac 2580
OY 2581 ttaactgctcctgac 2640
DB 2581 ttaactgctcctgac 2640
OY 2641 gcttacaac 2700
DB 2641 gcttacaac 2700
OY 2701 caaaacgctcctgac 2760
DB 2701 caaaacgctcctgac 2760
OY 2761 gttgaatcactgaac 2820
DB 2761 gttgaatcactgaac 2820
OY 2821 gcaagcagatgagcgttagctgtagcgaacacacacacacacacacacacacacac 2880
DB 2821 gcaagcagatgagcgttagctgtagcgaacacacacacacacacacacacacacac 2880
OY 2881 gtaactcactcactcactcactcactcactcactcactcactcactcactcactcact 2940
DB 2881 gtaactcactcactcactcactcactcactcactcactcactcactcactcactcact 2940

[illegible][illegible]

D5	421	CTAGAGCGCTGCTTCAAGAAAAAGTTTAGAAAAAGCAACAACTACTTCACAAAT	480
QY	481	ccctatgaaatlaaaacctcgacacgctcatttagatcgctalataygtlaaacaactcgt	540
D5	481	CCTTAGAATTAATAACCTGCGACAGTCATTTTAGATCGGTATATGGTTAAACAACCTGT	540
QY	541	gatttactcgcctcctcaatttaagaacaaagcacaagaactcgcgacacctaatttat	600
D5	541	GATTACTTCGCTGCTACATTTTAAGCAAAAGCAACAACTTCGCGACACCTTAATTTAT	600
QY	601	gatttaccgcttgacaaatgaagaacgacgagttacaagaacgtctgtaagaagcaattia	660
D5	601	GATTATTACGCTTGCAATGAAAGCGCGGAGTACAAAGCGCTGTGAAGAGCGCAATTTA	660
QY	661	gacaaagctaaagctcgtctgtgtgataatcaatcaacttaccataaagaataacagatgc	720
D5	661	GACAAAGCTAAAGCTCTGTGTGATCAATCAATCAATCAATCAATCAATCAATCAATGCT	720
QY	721	ttcaaaactgaaactaaacgaagaatagcgaaanaagcatatagtcagatgtaagctgcctt	780
D5	721	TTCAAAACCTGAAGCTAAACAAGATGAGGAAAAAGCATATGATCAGATGAAAGCTGCGCTT	780
QY	781	actcgaagaagttgaagaagttaagtgcatatcaacacacacaaacaaagctgtgtgattaa	840
D5	781	ACTCCAAAAGTTGAAAGGTAAAGTGCATTAACACCTCAAAAACAAAGCTGTGTAATTACA	840
QY	841	gcagtaaccagtgagcgaacacacaaatatacaacttcagctgctgcnaatgaaatata	900
D5	841	GCAATACCACTGACGCAACACTAAATTTACACTTTCAAGCTGCTGCAATGCAAGATACA	900
QY	901	gtanaagtaatactgtagcgtatctataaagtggaagtgtaaacatccatttgccctaat	960
D5	901	GTTAAACGTAAATACCTGTACGTATCTATTAAGTGGACGGTAACTTCATTTGCCCTTAAT	960
QY	961	acggacagctggttctctatacctacagacggaanaacatacactcgttgatgttcaactcca	1020
D5	961	ACGGACGAGTGTCTTTACTACTACGAGGAAACACTATCACTGTGATGCTTCAACTCCA	1020
QY	1021	ttcgaanaataatacgaagataaagaatgaatgaatgaatgaatgaatgaatgaatgaatga	1080
D5	1021	TTCGAAAAATTAATCGAAGATATAAGTAGTAGTTAAAGGCTTTAAAGCAAAAATGGCTAA	1080
QY	1081	gaatttaagaagaatgcatctcactctcaagcttcgaanaatgaatgctgtagttactcaagt	1140
D5	1081	GATTTTAAAGAAAGATGCATTCATCTTCAAGCTCGAAATGATGCTGTAGTTACTCAAGTG	1140
QY	1141	tttggaacaaatgtaacaaacaaacaaacttcgttaacttgacgacgagttacttcgcacact	1200
D5	1141	TTTGGAACAAATGTAACAACAAACAACTTGTAACTTACGACAGGTACTTTTCGACACT	1200
QY	1201	gacgaactttaaagctagtattgataagtggttgacacctgaagaactgtaaacagctcg	1260
D5	1201	GACGATACTTTAACAGTAAATATTGATTAAGTGTGTAGCACTTAACCTGTAACCACTTCG	1260
QY	1261	aacgltactatatacagaatgttgaaacatggaacaaacgcattccagtaattgcatctactct	1320
D5	1261	AACGTTACTATTACAGATGTGAAACTGGAACAAACGCAATTCAGTAATGTGACTACTCTCT	1320
QY	1321	ggttctacacaaatacattatcgttaaaagaagaaggtgtatcgtgtgtataacataataact	1380
D5	1321	GGTTCTACTAATTAATTAATTCGTTAAAGAAAGCGTTGTAGTACTGTTAAACAAATTAACCTT	1380
QY	1381	gctacaaataatgtaaaacatataactgtgttaacatgcaagaagcttaccagtttaagtgttc	1440
D5	1381	GCTATCAATTAATGTTAAACATTTAACTGCTTACAAATGCAAGAACCTTACGAGTTTAGTTC	1440
QY	1441	actgcaaaacgcaatcagcaccacacgtgtgtctacccgtcctactactttaggtgtgtataact	1500
D5	1441	ACTGC AAAAGCAACACACCACTGTTGCTACGCGTCTACTACTTAAAGTGTGTAACACT	1500
QY	1501	ttaactactggttctccttaccacaaatgttttggtgttaaatgtgctgtgtgtatgaatgaa	1560
D5	1501	TTATCTACTAGGTTCTCTTCAACAAATGTTTGCGTTAAATGCGTGTGTGTGGAATGAA	1560
QY	1561	gctggaactatatactcgtgtcttcaattcaacaaacaaacglttgctactaagtttagacgaa	1620
D5	1561	GCTGGAACCTATTATCTCGTCTCTCAATTCACAAACAGCTTGTGCTAAAGTTAGAGAA	1620
QY	1621	tctactcttagctgataactttgttatcttggttgaaagaacatctgtaagaatgttgtgtctc	1680
D5	1621	TCTACTTACTGATAACTTGTATTATTTAGTTGAAATAAATCTGTGTAAGTTGTGTCTCT	1680
QY	1681	gaactaaataataatgacagcgcacaaatggttaactttagtgccaagaacggaacttaaa	1740
D5	1681	GAACTAAATATATATGACAGCGCTAAATGTGAATCTTAACTGCCAAAAGCGACCTTAAA	1740
QY	1741	gaaataataatctatacaatacaataataaagaagcttgaaagttccgaataaagttatgaa	1800
D5	1741	GAAATATCAATCTATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT	1800
QY	1801	ttagcgaactgttaacgagaacaaatataatgatttcaaaactcaagaacttaactgtcctaca	1860
D5	1801	TTAAGCACTGTTTAACGAGAAACATATGAATCTCAAACTCAAGCTTAACCTGCTTACA	1860
QY	1861	gttattagcgttaaagctttaaagaagcgaagcttggaattaaagaatgaactggaagctcaaga	1920
D5	1861	GTTATTAGCTTAAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1920
QY	1921	tttactgtgaagctctcagaagaatttaaaatataatgatttcaaacggtttcgggttagc	1980
D5	1921	TTTACTGTGAAGTTCTCAGAGAAATTTAAATACATTTAACTGCTACAAACGTTTGGGTAGC	1980
QY	1981	acatacabaatcgtgtcaagttgctgtatgaagaacgagtgacaaactatctgctcttaca	2040
D5	1981	ACATACATACAGCTCAAGTGTGCTGTAGTAAAGCGGGCTCAACCTTAATGCTCTTACA	2040
QY	2041	gcaagtgacatcatcctcaagtaagtgttgaaagcggttacgtgtgaagtggaacatacaa	2100
D5	2041	GCAAGTGACATCATCTCAGCTAGTGTGTAAGCGGTACTGTCTCAAGATGGAACATTAACAA	2100
QY	2101	gtgaagctgtcgtcaacaaatagaacgaacgaacgaacgaacgaacgaacgaacgaacgaac	2160
D5	2101	GTTGAAGTCTGCTGTAAACCAATTAAGAACGTAACCAAGCTTAACAAATTAAGATGCTTGGT	2160
QY	2161	aaagtgcaacaaagctcctgtttaaagaatgctgcaaatgcaaatactttagaactaactat	2220
D5	2161	AAAGTGCAACAACTCCTGTTTAAAGATGCTGCAATCAATCAATCAATCAATCAATCAAT	2220
QY	2221	attatacaatttaacacttgaaagtgaaagaatgaacgacccaaaggtttacaaaaatgcttc	2280
D5	2221	ATCTATACATTTACATGTAAGGTGCAAGATGACAGTACACCAACGCTTACAAAAGTATTC	2280
QY	2281	aaagtgatattccttaaaagaacgctgtgacgttactacacttaacgaacgcttgatgaaggt	2340
D5	2281	AAAGTGATATTCTTTAAAGACGCTGATGCGATTACTACACTTAAGAACGTTGATGACAGGT	2340
QY	2341	caaaaaattcaattcaaatattgaacgaagaatlaaaaaactctagtggtttctttagtggt	2400
D5	2341	CAAAAAATTCACTATCCAAATTTACGAAGAAATTAAAAACCTCTAGTGGTCTTTAGTGGGT	24

QY 2641 cgttacacacgctgagacacacactgtagctctgtagaagctctgctgctgacaaagatgct 2700
 BSSBSA
 DB 2641 cgttacacacgctgagacacacacactgtagctctgtagaagctctgctgctgacaaagatgct 2700
 QY 2701 caaaagctctgctgtagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 2760
 DB 2701 caaaagctctgctgtagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 2760
 QY 2761 gttgaatcattgtagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 2820
 DB 2761 gttgaatcattgtagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 2820
 QY 2821 gcaagcagtagctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 2880
 DB 2821 gcaagcagtagctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 2880
 QY 2881 gttgtagctctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 2940
 DB 2881 gttgtagctctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 2940
 QY 2941 gttgtagctctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3000
 DB 2941 gttgtagctctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3000
 QY 3001 gcaagcagtagctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3060
 DB 3001 gcaagcagtagctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3060
 QY 3061 aaaaacattcaatgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3120
 DB 3061 aaaaacattcaatgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3120
 QY 3121 gttgtagctctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3180
 DB 3121 gttgtagctctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3180
 QY 3181 tacaacattgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3240
 DB 3181 tacaacattgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3240
 QY 3241 gtttcaaaagatttcaaaagatttcaaaagatttcaaaagatttcaaaagatttcaaaagatttcaaaagattt 3300
 DB 3241 gtttcaaaagatttcaaaagatttcaaaagatttcaaaagatttcaaaagatttcaaaagatttcaaaagattt 3300
 QY 3301 tctgttgcctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3360
 DB 3301 tctgttgcctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3360
 QY 3361 agctatcagtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3420
 DB 3361 agctatcagtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3420
 QY 3421 ccaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttccctt 3480
 DB 3421 ccaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttccctt 3480
 QY 3481 aatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttccctt 3540
 DB 3481 aatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttcccttcaatttccctt 3540
 QY 3541 gatgtagctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3600
 DB 3541 gatgtagctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3600
 QY 3601 aaaaacattcaatgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3660
 DB 3601 aaaaacattcaatgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3660
 QY 3661 gatgtagctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3687
 DB 3661 gatgtagctgtagctgtagaagctctgtagaagctctgtagaagctctgtagaagctctgtagaagctct 3687

RESULT 3
 LOCUS BSSBSA 3832 bp DNA BCT 25-OCT-1994
 DEFINITION B.stearothermophilus sbas gene for S-layer protein (SbsA).
 ACCESSION X71092.1
 VERSION X71092.1 GI:312729
 KEYWORDS cell envelope protein; cell membrane associated protein; S-layer protein; sbas gene.
 SOURCE Bacillus stearothermophilus.
 ORGANISM Bacillus stearothermophilus.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
 Bacillus.
 REFERENCE 1 (bases 1 to 3832)
 AUTHORS Kuen,B., Sleytr,U.B. and Lubitz,W.
 TITLE Sequence analysis of the sbas gene encoding the 130-kDa surface-layer protein of Bacillus stearothermophilus strain PV72
 JOURNAL Gene 145 (1), 115-120 (1994)
 MEDLINE 94320770
 REFERENCE 2 (bases 1 to 3832)
 AUTHORS Kuen,B.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-1993) Kuen,B., Institute of Microbiology and Genetics, University of Vienna, Dr. Bohrg. 9, 1030 Vienna, Austria
 COMMENT Related sequences: D90050 and Gilmore, R.D., Mol. Microbiol., 6:1579-1586 (1992).
 FEATURES
 source . 1..3832
 . Location/Qualifiers
 . /organism="Bacillus stearothermophilus"
 . /strain="PV72"
 . /db_xref="taxon:1422"
 . /gene="sbas"
 . 1..3767
 . /gene="sbas"
 . 37..3723
 . /gene="sbas"
 . /codon_start=1
 . /product="surface layer protein"
 . /protein_id="CA50409.1"
 . /db_xref="GI:312730"
 . /db_xref="SWISS-PROT:P35825"
 . /translation="MDRKAVALATASAASAFAANPNASEATDVAVVSQAKQ
 PKRAYTYSHTVETGEPPNINDVAYEYKAKRYDAVALYKAGAKADYADLO
 KEETVFEKANKSCARAVATYDAVATKIDEMOELEAVOALDERAKOYHRI
 PEIKRTVILDRVYKGTRODLRSFEKAKDELNDLSLYDITLANKAREVDYKAG
 NDKAKAADONQYIPKTYDAFKETLTPVAKKALDADPRALIPRESVSAINTOKA
 VELTAVPVGKTLQLSANEDVNAVTVRYKKGNTPPALNTADVSSTDGKIT
 VASTPEFNNTKRVVKGKIDKNGSEFEDAFTELRNDVAVTQVEGVNTNNTSVN
 LAAGTIDIDTLIVVDKLABETVNSVTLTVDVTEGRIPVIASTSGSTLTLEKE
 AVTGEKQYKIALINVTLLGVNAEAEVETANASAPVAVLPTLIGTSTLSITLT
 NYMGKLAGVNEAGTYPGLOFTTPPATLIDSLADNVLVEKESGVVAVSLKYNA
 DAKMTLVYKADLKENTIQKIKKGLKSKDGELGTVNEKYEEFTODTLA PTIVSY
 TSKNDAGLKTDAQEFYKESSENLENAVTVSGSTTYGVAVYKAKANSALTAS
 DIIIPASVEAVTQDGTKVKVANSLODEROGKRLVFGGATAPVADANANVTLAVY
 IYFTTEGQDVAPVTKVKKIDSLADNAGVTLLENVDAGORFTLQFSEELXTSSGSL
 VGGKATVEKTLNNGVNDAGTGVSAPEPDANGKVTAAVTLTGDRNDKDAKLIV
 VKSSSTDGIALDVAAGVYIKEDLLIRNSRNVYASVKAADGDGSAASAPSTVID
 TKSKLVERNEDIAVPEKENVYVDAAGNAYAGVTALDGSNTFVTPPSLEKAGT
 VTSVILIDVGRVGNNTSKYITSFKVSNAPLISLSDAGVAVNDKSTLTITERSDS
 VNPVITLLKAKGTSNTNTLVVNNKENTYILVEKGVTLDEFTQYELAVKQVGTG
 TDIDSKEVFTIGSVADBEVYPALVGVGSHNGSYQDADAETLRVADVAEVALQF
 SSGIDLNATVNTTDDKTEVEVTSKESVDADHDGAKETLVINTVPLVDNKT
 YKTVSGVMDAAGNADITTFIK"
 37..126
 /gene="sbas"
 127..3720
 /gene="sbas"
 /product="surface layer protein"
 3767..3806
 686 c 749 g 1042 t
 BASE COUNT 1355 a
 ORIGIN

Query Match 100.0%; Score 3687; DB 2; Length 3832;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgagtagaanaaaactgtgtaaacatgcaacagcaagtgctatgtcagcaagtgcaatt 60
DB 37 ATGCATGCAAAAAGAGCTGTGCAACTAGC AACCAGCAAGTGTATTGTCAGCAAGTGCATTT 96
QY 61 gtgcgtgaataatccaaagcttctgaagcggctacagatgtgtgcaaacgttagtaagccaa 120
DB 97 GTCCGTGCAAAATCCAAAGCTTCTGAAGCGCTACAGATGTGACACTAGTAAAGCCAA 156
QY 121 gcaaaagcagcttcanaaaagcacaatactatacttaacgcatatagtaacgaaactgt 180
DB 157 GCAAAAGCACTTCAAAAAGCATACATCTACTTACGCCATACGTAAACGAAACTGTGT 216
QY 181 gaattcccaaacatbaacgaatgtatatagtgtgaatacaacaaagcgaaaaaacyatacgt 240
DB 217 GAATTCGCAAAACATTAACGATGTATATGCTGATACAAACAAACGAAAAAGATACCGT 276
QY 261 gatgggtgagcttctgtgaataaagcaggtggtcggaaaaaagacgcttactatgtat 300
DB 277 GATCGGATAGCATTAAGTAAAGCAAGTGGCGGAAAAAGACCCCTTACTTACCTGAT 336
QY 301 ttacaaaagaataatgaacttaagcttttcaaaagcaaaccttaattctggcgaagctgt 360
DB 337 TTACAAAAGCAATATGAANAATACCTTTTCAAAAGCAAAACCTTAATCTGGCAAACTCGT 396
QY 361 gtagcaacttcatcgaatgtctatacaactatgaacaaatataagcaaatgtcgcaagag 420
DB 397 GTAGCAACTTACATGATGCTATACATCAACATGCAAC AAAATTAGACGAATGCGCAAGAG 456
QY 421 ctgagagctgtctgttcaagcaaaagatttagaaaaagcaagaacatactatacaaaatt 480
DB 457 CTAGAGGTGCTGTCAAGCAAGCAAAAGATTTAGAAAAAGCAAGCAATACATATACAAATT 516
QY 481 cctttagaatlaaaactcgcaacagctcatctttagatcggtatctgtgttaaaacaactcgt 540
DB 517 CCTTATGAATTTAAACTCGCACAGTCACTTTTGAATCCGATTAATGATTAACCAACTGCT 576
QY 541 gattactctgctctacatttaaaagcaaaagcacaagaactctgcagcagcttaattat 600
DB 577 GATTACTCTCCCTCTACATTTTAAAGCAAAAGCACAAAGAACTCCGACAGCTTAATTTAT 636
QY 601 gatattacgtttgcataagaagcgcggaaggtacaaagacgctgtgtaaaagcagcaattta 660
DB 637 GATATTACCGTTGCATATAAAGCGCGGCAAGTACAAAGACCTGTGAAACGACGCAATTTA 696
QY 661 gacaagctlaaagctgtctgttgaataaatcaatcaatacttaacaaagtaacagatgt 720
DB 697 CACAAACCTTAAAGCTGCTGTTCATTAATCAATCAATCAATTTCCAAAGTAACAAATGCT 756
QY 721 ttcaaaactgaactacacagaagtagcgaaaaaagcattatagatgtcagatgaagctgcgt 780
DB 757 TTCAAAACTGAACCTAACGAAGTAGCGAAAAAGCATTTAGATGAGATGAACCTCGGCTT 816
QY 781 actccaaaagtgtgaaggtgaatgtgcgatttaacactcaaaacaaagctgtgttaattaca 840
DB 817 ACTCCAAAAGTGAAGTGTAAAGTGCATTAACACTCAAAAACAAAGCTGTTTAATTAACA 876
QY 841 gcaagtacagagtaaacggaacactaaatttaacactttcagctgtcgcaaatgaatgaatca 900
DB 877 CCAAGTACCAAGTGAACGGAACCTTAATTAACACTTTTCAGCTGCAATATGAATGAATCA 936
QY 901 gtaaacgtlaaactgtatctataaagtgtgaacggttaacactccatttgccttaatt 960
DB 937 GTAAACGTAATAATGCTGTATTAATTAATGAAGTGAACGGAATCAATTCATTCCTTAAT 996
QY 961 acggagagatgttcttctatctacagaacggaaaacatactatcgtgtgaatcctaactcca 1020
DB 997 ACGCAGATGTTCTTTATCTACAGACGAAAAAATATCACTGTGTGATGCTTCAACTCCA 1056

QY 1021 ttccgaataatcaggaatataaagtagtaaggtatataaggtataagcaaaaatgtgcaaa 1080
DB 1057 TTCCGAATAATATAGCAATATAAGTATAGTAAAGGTATTAAGCAAAAAATGGCAAA 1116
QY 1081 gaatttaagaagatgtacttaacttcaagcttcgaaatgtatgtgttacttaagtg 1140
DB 1117 GAATTTAAAGAGATGCTATTCACCTTTCACCTTCGAAAGTATGCTGTATTCATCCTAAAGTG 1176
QY 1141 ttggaaactaatgttaacaaacacactctgttaaacttagcagcaggtatcttgacact 1200
DB 1177 TTGGAACTTAATGTAAACAAACACCTCTGTAACTTAGACGCAAGTATCTTCACACT 1236
QY 1201 gacgaacttaacagtagtattgttaagtgtttagaacctgaaactgttaaacagctcg 1260
DB 1237 GACGATATTTAACGATGCTATTTGATTAAGTGTTAGACACCTGAACCTGAACGCTCG 1296
QY 1261 aacgttactatbacagatgttgaacgttgaaacagcattccaglaattgtcatcttct 1320
DB 1297 AACGTACTATTACAGATGTGAACCTGAAACGCATTCAGATTAATGCAATCTACTTCT 1356
QY 1321 ggtttacaattactattacgttttaaaagaagcgttagaactgtgtaaacataataact 1380
DB 1357 GGTCTTACAAATTACTATTACGTTAAAGACGTTAGTAACTGTTAAACAAATATAACTT 1416
QY 1381 gctatcaataatgttaaaacataactatgtttacaatgacgaagcttaagtagtctc 1440
DB 1417 CCTATCAATTAATGTTAAACATTAACCTGTACAAATGAGAAAGCTTAGAGTACTGTTC 1476
QY 1441 actgcaaacgcatacgaacccaactgttgcacgcgtcactactttagtgtgtaacaact 1500
DB 1477 ACTGCAACGCAATGACGACCACTGCTTCTACCGCTCTACTACTTTAGTGGTCAACT 1536
QY 1501 ttatctactgttctctttaaacaacatgtttgggttaattgtgctgtgtgtgtgaatga 1560
DB 1537 TTATCTACTGATCTCTTACCAAAATGTTGGGTAAATGGCTGTGTGTGAATGA 1596
QY 1561 gctggaacttatacctcgtgcttcaacttcaacaacaaagcttgcataagtttagacga 1624
DB 1597 GCTGGAACCTTAATTCCTGCTGCTTCAATTCACAAACAGTTTGCATTAATGACGCA 1656
QY 1621 tctacttagctgataaactttgtaattagttgaanaagaaactgtgaacgttgtgtctct 1680
DB 1657 TCTACTTAGCTGATTAACCTTGTATAGTTGAAAAAGCAATCTGATAGTGTGCTCTCT 1716
QY 1681 gaacataaatttaatgacgaagcttaaaatgttaacttaagtgccaaagcgaacttaaa 1740
DB 1717 GAACATAATATAATGACGACGCTAAATGTAACTTAATGACCAAAAGCGGACCTTAAA 1776
QY 1741 gaaatataactatcaatcaaatcaaaatgaagagcttgaagtcgataaaggtatga 1800
DB 1777 GAAATTAACATCTATCAAAATCAAAATTAATAAAGCTTGAAGTCCGATTAAGGTGCA 1836
QY 1801 ttgaagcactgttaacgagaaacatatagttcaaaaactcaagacttaactgtctctaca 1860
DB 1837 TTAGCACTGTTAAACGAAAAACATATGATGTTCAAAAATCAAGACTTAAGCTCTCA 1896
QY 1861 gttattagcgaagcttcaaaaatgtgcagcgtgtgataaaagtatacgtgaagctcaaaa 1920
DB 1897 GTTATTACCGTAAGCTTCAAAAATGGCGACCTGATTAAGTAAGTGAAGCTCAAAA 1956
QY 1921 tttaactgtgaagttctgaagaatttaatacttaacttaagtatacgaacgctttcggtgac 1980
DB 1957 TTTACTGTGAAGTCTTCGAGAAATTTAAATTCATTTAAATGCTTACAAACGTTTCGGGTAC 2016
QY 1961 acaatcacatcaggtcgaagctgtctgttagtaaaagcgggtgcaaaccttaactgtctctaca 2040
DB 2017 ACAATCACATACGGCAAGTGTGTAATTAAGCGGGTCAAACTATCTGCTTACA 2076
QY 2041 gcaagtgacatcattccagcctagtgttgaagcgggttaactgtgtcaagaatgaacatacaaa 2100
DB 2077 GCAACTGACATCATTCACGCTAGTGTGTAACCGGTATCTGCTCAAGATGGAACATACAAA 2136
QY 2101 gtgaagatgtctgttaaccaaattgaagacgttaaccaaggttaacaaattagtagtctggt 2160

Db	2137	GTGAAGTTGCTGCTAACCAATTAGAACGTAACCAAGGGTACAAATTAGTAGTGTCCGT	2196
Oy	2161	aaagtgcaacaggtccctgttaagaatgtctgcaaatgcaataactttagcaactaact	2220
Db	2197	AAAGGTGCAGACGCTCTGTTAAAGATGCGCAAAAGCAAAATCTTAGCAACTACAT	2256
Oy	2221	atctatcaacttaeaactgaaggtcaagcgttaacagcacaaggttcaacaagtatct	2280
Db	2257	ATCTATACATTTTACAACTGGAAAGGTCAAGAGTATACAGCCAAACGGTTACAAATATCTC	2316
Oy	2281	aaagttactctttaaagaagcgcctgactgaattacatacctaagcaagttgagtcaggt	2340
Db	2317	AAAGGTGATTCTTTAAAGACGCTGATGCAGTTACTACCTTACGAAGTGTATGCAAGT	2376
Oy	2341	caaaatctcaactlccaaattlaagcgaagaattaaaactctagttgtctttaagtt	2400
Db	2377	CAAAATTCACATATCCAAATTTAGGGAAGATTTAAATCTTGAGTGGTTCTTAGTGGGT	2436
Oy	2401	ggcgaagtaactgtcgagaataatcaacaacacgtgttgatgtcgtactggaaca	2460
Db	2437	GGCAAAAGTAACTGTGCGGAATTTACAAACAAAGGATGGGTGAGAGTGGACTGGAAACA	2496
Oy	2461	acgtatcaagttgtccttaagacagatgcaaatgtrtaagtaacagcgcgcgtgttaca	2520
Db	2487	ACTGTATCAGTTGCTCCTTAAGACAGATGCAATAGGTAAAGTATACAGCTGTGTGGTTACA	2556
Oy	2521	ttaactgtgtctgacaaatacgaacaagaatgcaaatltgcgtctgttagtataagct	2580
Db	2557	TTAACTGGTCTTGACATTAACGACAAAGATGCAATTCGCTGTGGTATGATATAGCT	2616
Oy	2581	tctactgattgaatttgtctgtgtatgttaataagaaaaagatactttaatt	2640
Db	2617	TCTACTGATGGAATTCGTGATGTGAAGCTGTAAGTAAATTAAGGAAAAAGATATTTAATT	2676
Oy	2641	cgttaacaaagcttgagacacacactgaagctctctgtgaagcgcgcgcgcgaacaagatggt	2700
Db	2677	CGTTATCACAGCTGTGAGACACACTGTATGCTTCTGTAAAGCTGTCTGTACAAAGATGCT	2736
Oy	2701	caaaagcctctgtctgcatctcccaacaagaacatgcaattgatacaactaaagattatla	2760
Db	2737	CAAAAGCCTTCTGCTGATTCCTCCACAAAGACATGCAATTTGATGACAACTAAGAGCTTATTA	2796
Oy	2761	gtctaatcaatgaactgtatcttaagcggaaatgaactggaacaacatcgttgttaagaat	2820
Db	2797	GTTAAATTCATGTAACTGTGATTTTAGGGGAAGTTAACTCGAAMACATGCTGTGTAAAGAT	2856
Oy	2821	gcaacaggttaatgaggtagctgtgtacgttaacagcaattagaaggtctcaacaataaatt	2880
Db	2857	GCACAGGTTAATGCGGTAAGCTGCTACTGTATACAGCAATTAGAGGTTCTTACAAATTAATTT	2916
Oy	2881	gtattcactcaactccaagaattaaaagctgtagcagtttactctgttaacaaatgacgt	2940
Db	2917	GTAATTCACCTCATCTCAAGAAATTAAGAGCGGACAGTTTACTCTGTAAACATTAAGCGT	2976
Oy	2941	gtgagagataaagttagttaacacaactcttaataacttaactcgtttaagaagctatct	3000
Db	2977	GTGAGAGATTAAGTAAGGTATACACAACCTCTTAATATATTACTTCGTTCAAGACTATCT	3036
Oy	3001	ggcaatccaaggttatcttaactcaagcatgtgcagaggtgaggttaagttgaacggttct	3060
Db	3037	GGCAATTCAGGTTATCTTCAATCAGATTGCTGACGGGTGACATTAAGTTACGTTACCGTCTT	3096
Oy	3061	aaaaaatataaatgttaattcaagcagattcaagttccaaaccccaacatcacctttaagag	3120
Db	3097	AAAACAAATTACAAATTGAATTCAGGATTCAGTTCCAAACCCAAACATCACTTTTAAGAG	3156
Oy	3121	gtctagcgaacttcaattactaaattcacacttagtaaatgttaataatgnaataaaca	3180
Db	3157	GCTGACGGAATCTCATTTACTAATTACACTTTAGTAAAGTAAATATAGAAAAATTAACA	3216
Oy	3181	tacaaaattgtatccacaagaagtgtaaaccttgacaggttactcaataatgattagca	3240

D6	3217	TACAAATGTGATTCCACCAAGAGGTGTAACACCTTGAGAGTTTACTCAATATGATGATTACCA	3276
Q7	3241	gttccaaaagattttccaaactggtactgatatgtagaagaaagttacattcatctccaggt	3300
D6	3277	gTTTCAAAAGATTTCCTCAAGCTGGTACTGATTTGATAGCAAGTTACATCTATCCAGGT	3336
Q7	3301	tctgttgctatgacgagagttaaacctgctcagtagagcgttggatctca tgsatggaaca	3360
D6	3337	TCTGTGCTATAGACGAAGTAATAACTGCTGCTAGTAGAGGCGTTGCTTCATGGAATGGAAACA	3396
Q7	3361	agctatactcagagatgctgcagcgaacacgagctcggctcgtagctgactcgttgcgag	3420
D6	3397	AGCAATACTCAGGATGGTGGACGACAAACGACATCTGGTCTGTAGCGACATCTGTTCCGAG	3456
Q7	3421	ccattgcacctcaattctcagaaggtatcgatttgaagaaatgcaactgttgcagtaaca	3480
D6	3457	CGAGTGGCCTTCATTTCTCAGAAAGTATGCAATTTAAGCAATGCAACATGTGACATAACA	3516
Q7	3481	aataatactgatgtaaaacctgtgagttattttccaaaagagtgtagacgagacat	3540
D6	3517	AATATTACTGATGATGAATAAATTGTTGAAGTATTATTTCAAAAGACAGTGTAGACCAACCAT	3576
Q7	3541	gatgcaggtgtgtactaaaggagacattaaagtaattaaacacgttaacccattgtaattgat	3600
D6	3577	GATCCAGGTGCTATTAAGGAGACATTAGTAATTAACAGTTACTCTTTAATATTTAT	3636
Q7	3601	aacgcgaagaatttaaaagatctgtglaagtggaatgaaagatgcagcaggtatgtttgca	3660
D6	3637	AACGACAAAGCTATTAAGATTTGTTGAAGGAGATTAAAGATGACAGAGTAAATGTCCA	3696
Q7	3661	gatactattacattctctataaagtca	3687
D6	3697	GATACATACTCATTTCTAATTAAGTAA	3723

```

RESULT 4
AF055578 LOCUS 3680 bp DNA BCT 08-MAR-2000
DEFINITION Bacillus stearothermophilus surface layer protein precursor (sbpc)
gene, complete cds.
ACCESSION AF055578
VERSION 1
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Jarosch,M., Egelseer,E.M., Mattanovich,D., Sleytr,U.B. and Sara,M.
TITLE S-layer gene sbpc of Bacillus stearothermophilus ATCC 1980:
molecular characterization and heterologous expression in
Escherichia coli
JOURNAL Microbiology 146 (Pt 2), 273-281 (2000)
MEDLINE 20170659.
REFERENCE
AUTHORS Jarosch,M., Egelseer,E.M., Mattanovich,D., Sleytr,U.B. and Sara,M.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-1998) Zentrum fuer Ultrastrukturforschung,
Universitaet fuer Bodenkultur, Gregor Mendelstrasse 33, A-1180
Vienna, Austria
REFERENCE
AUTHORS Jarosch,M., Egelseer,E.M., Mattanovich,D., Sleytr,U.B. and Sara,M.
TITLE Direct Submission
JOURNAL Submitted (13-APR-1999) Zentrum fuer Ultrastrukturforschung,
Universitaet fuer Bodenkultur, Gregor Mendelstrasse 33, A-1180
Vienna, Austria
REMARK Sequence update by Submitter
COMMENT On Apr 13, 1999 this sequence version replaced gi:3025825.
FEATURES
source
..
/organism="Bacillus stearothermophilus"
/strain="ATCC12980"
/db_xref="taxon:1422"

```



```

mRNA      /db_xref="ATCC:12980"
           51..3643
           /gene="sbpc"
gene        51..3643
           /gene="sbpc"
misc_feature 84
           /gene="sbpc"
           /note="alternative transcription start site"
RBS        255..261
           /gene="sbpc"
sig_peptide 273..362
           /gene="sbpc"
           /gene="sbpc"
           /note="S-layer protein; sbpc"
           /codon_start=1
           /transl_table=11
           /product="surface layer protein precursor"
           /protein_id="AAC12757.1"
           /db_xref="GI:3025826"
CDS
           /translation="MDKKAVRLATASAVASAFVAPNPHASQANTVATVYQAKAO
           MKEAYTTHSTTEGPPDIDYTAAYKAKQARNAAYNKNAGAKKADYLDLQ
           ALTEYVRANKSGEARYATYIDATNATKIDKMKQELAAVADADLKABEYLHI
           SYELKTRIVLIDRVYGOSTRRELIRSTFKADQALRDRLDYDTVAKAREADAVNAG
           NIDKRAALDQVNOYVSKYTDFAKELQKADQAKAYAPALPKVESADISTSEK
           VYFKPVDRKATLIPKNSITLKGTEKLPKSEVESGLFATVTLIDYDLGKRTYV
           VTSGLKDRPKRETELSTNEFTYKPPASITENNELPDSAVDLKRYTVDDAGNV
           IKSFEELETSEKLTQCKFTYTGKASTVATYKGVTVTGGNYTLAVEDKAEVS
           ELKLRKKNVEVTLTANGNAFDKDSGLISSGLTLTAKKDOIENLTIKRVAGIDTFE
           ESLNBEVLVAVDGSVTPVPTALVKYGEVTKIPIVTVANPLYETLIVDSGVN
           VAKGKAEKFKLKDQYGNKFTGNVNTSDKTEITAVSNGSIGSEETVYVNGVAE
           GSTTILIKSGTEKEVYVAVGAPVANOIKYLDDEKIDKSAIESPANDVOLAKYA
           VDAKNVIGDITINDVTLTSEADTNGVYNAKSTANDPVTIVTINDSGKVKGETIT
           VKGTVTLGTVNVEVITLTKATVYTKRDKDLIELHADAQDGLAKILNEDLIDKQNG
           PVDNSAFTNTEKIQALNSVLSGIYSSPTSVIGSVNDMLDQDSISGLVAKAGT
           WTLTIVENEDSKLAPLAIYVKAAPAQDGVTVGLDVGAVGVGKTKTADKLSG
           HRLYAVDSDAVPAVATGRNSTKEANETVGTETVANAQOITIVIVDSDDRVAG
           YRTVEADLISVADKGTGTATVTPGNGOVTGKTLAVSDLANGHLYAAAGSS
           AAAAPVKGIAYDTVYRTYGTVEVTSGLTEVVAQDSQHLSIIEVENGVGKDTTI
           TGIQITKESAS"
mat_peptide 363..3569
           /gene="sbpc"
           /product="surface layer protein"
terminator 3606..3643
           /gene="sbpc"
BASE COUNT 1265 a 694 c 766 g 955 t
ORIGIN
Query Match 13.9%; Score 513; DB 1; Length 3680;
Best Local Similarity 66.6%; Pred. No. 3e-75;
Matches 767; Conservative 0; Mismatches 375; Indels 9; Gaps 2;
QY 1 atggatgggaaaaaagctgtaaacctagcacagcgaagtcatttcagcagaatgcattt 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 273 ATGAGCAAAAGAAAGAGAGCAAGTCGCACAGCCGCTGTAGAGAGAGCGCTTTC 332
QY 61 gtgcgtgcaaaatcccaacgcttcctgaagcggtcagatgtacgaacagtagtaagccaa 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 333 GTGCGAGCCCAACCCGCTGCTCCACAGCAAGCAAGCACTGCGAGCGTCTGACGCCAA 392
QY 121 gcaaaaggcagcgttcaaaaaatactactactacagcgcatacagtaacgaaacgtgt 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 393 GCGAAAGCGCAAAATGAAGAGCAATCTACTACAGCCATACGGTAACGGAAAGCGGC 452
QY 181 gaattcccaaacattaaagcgtatgtatgtcgtgaatacacacaaacgaaaaaagataccgt 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 453 GAATTCGCCGACATTAAGATGTATATAGCGCTACACAAACAAACAAAGGTATAGCA 512
QY 241 gatgcgttagcattagtgaataaagcaggtgcgcgcaaaaaaagcgttacttagtgat 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 513 AACGCGTGTGCTGCTGAACAAGCCGCGCGCAAAAAAGACGCGGTATTTGGCTGAT 572
```

```

QY 301 ttacaaaagaatatgaacttaccgttttcaaaagcaaaccttaccgttcggaagctcgt 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 573 TTGCAAGCTATATATAGCAAAAGTATGTATTC AAGCCAACCCGAATACGGCGGAAGCGCG 632
QY 361 gtatgaacttaccatgctgcttacaactatgcacaaaatttagcgaattgcgcgaag 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 633 GTGGCTACGATCAATCGACGGCTTACAACTATGCGACAAAATTAACAAAATGGCCAAAGA 692
QY 421 ctgaagcctgcgtgttcaagcaaaagatttagaaaagcagaacatactatcacaaact 480
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 693 CTCAAGCAACCCCTGCATGCGAAGATCTGAMAAAAGCAGAAAGCTCTTACCAAAAATT 752
QY 481 cctatgaataaataaactcgcacagtcattttagatcgcgtatagtgttaaaacactcgt 540
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 753 TCTTACGAGCTCAAAAAGCGGACGCTATCTGTGACCGGTCATAGCGCAATGACGCGG 812
QY 541 gattccttcgctctacattttaaagcaaaagcacaagaactctggagcgtctaattat 600
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 813 GAACCTCTCGCTCGACAGTGCACAAAGCCGATGCTCAAGCATTTGGCGACACGATTGATCTAC 872
QY 601 gatattacgcttgcgaatgaagcgcggaagctacaagaacgctgtgaagcaggaattta 660
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 873 GACATCACGCTTCCATGAAAGCGCTGAAAGCGCAGGACAGCTCTCAAGCCGCACTTG 932
QY 661 gacaaagctaaagctgctgttgaatcaatcaatcaatcttaccacaaagtaacagatgct 720
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 933 GACAAAGCGCAAAAGCCGCGCTGACCAAGTACCAATACCTATGCAAGTAAGCAATGCG 992
QY 721 ttcaaaactgaactaacagaagtagcgaataaagcatttagatgcagatgaagctgcgtc 780
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 993 TTCAAGCGCAAACTGCACAAAGCAGCAAGACGCCAAGCGCGCTATGAGCAGCATG 1052
QY 781 accccaaagcttgaaggttgaagtgatgtgaattaaaccccaaaagacgttgaattaa 840
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1053 ACCCGAAGGTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1112
QY 841 gcaatcaccgtgcagcaagcaactaaatactacaacttcagctgcgtcaaatgaagata 900
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1113 AACCGTGTAGCAAGCAAGCAAGCAATTCGTAAGAACTTCCTATTCCTTAAAGAGTACT 1172
QY 901 glaaacglaaatacgtacgtatcctataaagtgagacgtaaacattcattgccttaac 960
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1173 GAATA---CTAAACTTATCTCCTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGT 1229
QY 961 acgycagatgcttctcctatctacagacggaataactacatcgtgtgatgttcaactcca 1020
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1230 GTAACCTTTATATACATTAGTAGACGGTAAACTTACACTGTTTAACTATGTGT--- 1286
QY 1021 ttgcgaataatacggagataaagtagtagttaaagataaagacaaataagcga 1080
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1287 ---TTAAAGATTCACCTGTGTAAGATTTGAACCTAGCAACAGCAATTCACCTTACAC 1343
QY 1081 gaattaaagagatgcatcaccattcctcaagcttcgaataatgctgttagttactcaagt 1140
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1344 AAGCCAGTACACACTCTCTATACATTCACACTCAATAAAGTATACCTGAAGATAGTGCAT 1403
QY 1141 ttgtgaactaa 1151
   |||||
Db 1404 GATTTGACTTA 1414
```

RESULT 5
AF228338 3090 bp DNA BCT 24-JUL-2000
LOCUS Bacillus stearothermophilus Sbsd precursor (sbsd) gene, complete
DEFINITION
cgs.
ACCESSION AF228338
VERSION AF228338.2 GI:9188638
KEYWORDS
SOURCE Bacillus stearothermophilus.
ORGANISM Bacillus stearothermophilus
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.


```

CC FT /note="malose-binding protein"
CC FT 1150..1174
CC FT /note="ECOR1 adaptor"
CC FT 1174..1195
CC FT /note="immunogenicity protein"
CC FT 1194..1218
CC FT /note="ECOR1 adaptor"
CC FT 1219..1239
CC FT /note="sequence derived from pMAL-c vector"
CC FT
XX
FH Key Location/Qualifiers
FH 1..3399
FT /db_xref="taxon:32644"
FT /organism="unidentified"
FT
XX Sequence 3399 bp: 1577 A; 509 C; 797 G; 516 T; 0 other;

```

```

Query Match 2.1%; Score 77.4; DB 22; Length 3399;
Best Local Similarity 43.0%; Pred. No. 0.0014;
Matches 436; Conservative 0; Mismatches 576; Indels 3; Gaps 1;

QY 84 tgaagcgcgtacagatgtagcaacagtagtaagcgaagcgaagcgttcaaaagc 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2064 TGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2123

QY 144 atactatcttaacagcctacagtaagcgaagcgttgaattcccaaatlaagcgt 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2124 AATACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2183

QY 204 atatgctaatcaacaaagcgaagcgaagcgttgaattcccaaatlaagcgt 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2184 AGAAGTAATACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2243

QY 264 agcaggtgcgcgaagcgaagcgttgaattcccaaatlaagcgt 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2244 AACACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2303

QY 324 cgtttcaagcgaagcgaagcgttgaattcccaaatlaagcgt 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2304 TGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2363

QY 384 caactatcaacaaagcgaagcgaagcgttgaattcccaaatlaagcgt 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2364 AGAAGTAATACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2423

QY 444 agatttagaagcgaagcgaagcgttgaattcccaaatlaagcgt 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2424 AGAAGTAATACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2483

QY 504 agtcatcttagatcgctatgtgaagcgaagcgttgaattcccaaatlaagcgt 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2484 AACACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2543

QY 564 agcgaagcgaagcgaagcgttgaattcccaaatlaagcgt 623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2544 AGAAGTAATACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2603

QY 624 gcgcgaagcgaagcgttgaagcgaagcgttgaattcccaaatlaagcgt 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2604 AGAAGTAATACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2663

QY 684 tcaatcatcaatcttcaagcgaagcgttgaattcccaaatlaagcgt 743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2664 AGTAACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2720

QY 744 agcgaagcgaagcgttgaagcgaagcgttgaattcccaaatlaagcgt 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2724 TGAAGTAATACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2780

QY 804 tgcgattcaacatcaagcgttgaattcccaaatlaagcgt 863
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 2781 AGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2840
QY 864 aaattacaattcagcgttgaagcgaagcgttgaattcccaaatlaagcgt 923
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2841 AGAAGTAATACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2900
QY 924 ctataagtgagcgttgaattcccaattgaccttaagcgaagcgttccctac 983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2901 AACACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2960
QY 984 agagcgaagcgaagcgttgaattcccaattgaccttaagcgaagcgttccctac 1043
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2961 AGAAGTAATACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 3020
QY 1044 agtgaattgaagcgttgaagcgaagcgttgaattcccaaatlaagcgt 1098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3021 AAAAGTAATACATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 3075

```

```

RESULT 8
AP002460 111554 bp DNA PIN 19-JUL-2000
LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone:RID9.
DEFINITION AP002460
ACCESSION AP002460
VERSION AP002460.1 GI:8347620
KEYWORDS
SOURCE

```

```

ORGANISM
Arabidopsis thaliana (strain:Colombia) DNA, clone_11b:IGF BAC
clone:RID9.
Arabidopsis thaliana
Bakaryova, Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
REFERENCE
Nakamura, Y.
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. III
JOURNAL Unpublished (2000)
REFERENCE
2 (bases 1 to 111554)
AUTHORS Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2000) to the DDBJ/EMBL/Genbank databases.
Yasukazu Nakamura, Kazuo DNA Research Institute, Department of
Plant Gene Research, 1532-3, Yana, Kiserazu, Chiba 292-0812, Japan
(E-mail: yna@kamu.kazuo.or.jp, Tel: 81-438-52-3935,
Fax: 81-438-52-3934)
COMMENT
Addresses for correspondence: kase@kazuo.or.jp
For the latest information on annotation of this clone, please see
http://www.kazuo.or.jp/kase/cgi-bin/sgd.graph.cgi?c=RID9
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Graal
(informatics group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremlin.zool.lastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Gean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is T889 and the 3' clone is T789.

```

FEATURES

SOURCE

```

1..111534
/organism="Arabidopsis thaliana"
/strain="Colombia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="RID9"
/clone_11b="IGF BAC"
complement(join(4747..5195,5297..5754,5827..6284,

```

CDS

```
6457..6803,6864..7079,7216..7286,7326..7810,7914..8253,
8311..8737,9083..9347)
/feature="gene_id:FID9.1"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="helicase-like protein"
join(11589..11736,11836..12019,12305..12497,12774..12896,
13021..13080)
/feature="gene_id:FID9.2"
/codon_start=1
/evidence-not_experimental
/product="ATP-dependent CLPB protein-like"
/protein_id="BAA97091.1"
/db_xref="GI:877573"
/translation="MLQDLHERVLAQDLYESVADATRCSCAGISDPNRIASFMFK
GPIVSQVGVGVYGVYEDGKSTEVRRRPYSVQVDEIKPHRPVSIQLQLD
DGRITDSGSLHILLETIRNEDIKEAVYEMKQVVELARKTPKPMNIDIVLS
QPLNSVSKIVELQVQVQMRQKRLDEQPNRKLEITLVLKQDPAENDTILIDADQPF
NKLVIKKNENAHYKON"
complement(join(19018..19216,19317..19390,19600..19746))
/feature="contains similarity to unknown protein"
/feature_id="FID9.3"
/codon_start=1
/evidence-not_experimental
/protein_id="BAA97092.1"
/db_xref="GI:877574"
/translation="MAKERKRVGVREEDYKRMKILIMPLNSLHQALRSGYF
STLKITNYKELEKMKIRQEFVASFSSGFSFGAYIPCAKILINDFGRHGLAVL
PGKILITLVSKSSSKIGIKDKLIIIVIVKLVMP"
join(21817..22004,22075..22143)
/feature="contains similarity to ATP-dependent CLPB protein"
/feature_id="FID9.4"
/codon_start=1
/pseudo
/evidence-not_experimental
complement(22489..23812)
/feature="contains similarity to retroelement pol
polyprotein"
/feature_id="FID9.5"
/codon_start=1
/pseudo
/evidence-not_experimental
join(24404..24543,24659..24879)
/feature="contains similarity to ATP-dependent CLPB protein"
/feature_id="FID9.6"
/codon_start=1
/pseudo
/evidence-not_experimental
complement(25073..29379)
/feature="gene_id:FID9.7"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="retroelement pol"
join(30097..30190,30265..30551,30614..31510,31758..32036)
/feature="gene_id:FID9.8"
/codon_start=1
/pseudo
/evidence-not_experimental
join(40639..40882,41779..42326,42393..42794,42903..42936,
43020..43142,43226..43333,43425..43481,43588..43650,
43733..43973,44026..44107,44503..44674,44813..44940,
45043..45198)
/feature="gb|AA025852.1"
/feature_id="FID9.9"
/feature="strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
join(46889..47780,47984..48332,48695..48783)
/feature="gene_id:FID9.10"
/feature_id="FID9.10"
/feature="strong similarity to unknown protein"
/feature_id="FID9.11"
/codon_start=1
/pseudo
/evidence-not_experimental
join(51111..51234,51470..52398,52474..52672,52762..53360,
53433..53558,53717..53809,53908..54010,54110..54210,
54472..54531,54627..54728,54795..55047,55142..55226,
55334..55637,55730..57389,57601..58187)
/feature="gene_id:FID9.11"
/feature_id="FID9.11"
/feature="strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
complement(join(58782..58995,59066..59150,59231..59498,
60276..61922,61970..62259,62305..62482))
/feature="gene_id:FID9.12"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="maize transposon MudR protein-like"
complement(join(63866..64008,64160..64226,64346..64414,
64504..64620,64743..64790,64956..65053,65183..65264,
65486..65566,65677..65718,65854..65952,66106..66144))
/feature="contains similarity to human diaphanous 1"
/feature_id="FID9.13"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01984.1"
/db_xref="GI:9294133"
/translation="MLSKVKLPIDLSVYALDELVSVDNDVNLKQPTKEAEI
LKGFTGNKETIGRCQEPFLDLAKPYREIKRNVSEFKIQHSQVSGSKKRINQITL
SLGNALHGTARSSAIGFLDSILKLTDETRKRVSKTLMLHLYLGGVLAERKPLGLNP
KDMTQLVLAEMEQATSKGLEKVVQETASITDQISHEFKNKETLSVAGGVRSI
ASLTIVGSGSDALALYEGEDPARVPEQVSTLQNFVIRIVRSHHEENCKQVEERKR
AOKAENKELKGYNNEN"
complement(66524..66974)
/feature="gb|AA023008.1"
/feature_id="FID9.14"
/feature="similar to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
complement(join(68701..69561,69697..70591,70640..70667,
71526..72709,72898..73300,73370..73533,73687..74594))
/feature="gene_id:FID9.15"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="retroelement pol"
complement(75178..76058)
/feature="gb|AA023008.1"
/feature_id="FID9.16"
/feature="similar to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
76515..80547
/feature="emb|CA875469.1"
/feature_id="FID9.17"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="copia-type reverse transcriptase-like protein"
complement(join(81294..81521,81620..82285))
/feature="gb|AA023008.1"
/feature_id="FID9.18"
/feature="strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
```



```

OY 341 ctzaatctggcgaagctcgttagcaactacatcgaatcgaactaagcaaat 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3006 AAGAGTAAAGAAAAGATACGAAAGCAAGATTAAGAGATAGAACAGAAAAGAAA 3665
OY 401 tgaagcaatgcgcgaagcttagagctgcgttcaagcaaaagattagaagaagcg 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3066 AAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3125
OY 461 aacacat---acatcaccaaatcccttaagaattaaactgcacagcatttgatc 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3126 AAGAAATTAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3185
OY 518 gctatattgtaaaacaaactcgtattacttgcctcatttaagcaaaagcaag 577
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3186 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3445
OY 578 aactcgcgcgaagcttaattatgataatcgttcgaatgaagcgcggaagtaacag 637
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3246 AAGATTAACCCCAAAACAAAGATTAAGGTATAGAACAGAAAAGAAAAGAAAAGTAAAG 3305
OY 638 agcgtctgtaagcagcgaatttagacaagctaaagctgctgttagtaaatcaatcaat 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3306 AAAAAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3365
OY 698 acttaccaaaagtaacagatgcttcaaaactgaactaagcagaagtagcgaanaagcat 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3366 AAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3425
OY 758 tagatgcagatgaagctgcgttaactccaaagctgaagtgtaagtgatgaactacac 817
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3426 AAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3485
OY 818 aaaaaaagctgttaattacaagcagtaacagcgaagcgaactaaataatacaactt 877
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3486 AAGAAATTAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3545
OY 878 cagctgcgcgaatgaagctacagtaacgaactgaactgaactgaactgaactgaactga 937
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3546 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3605
OY 938 gtaacatcattcgtcccttaatacagcgaagtgcttcttattacagacggaagaaacta 997
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3606 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3665
OY 998 tcaactggtatgcttcaactcattcgaataataatcaggaagtaagtagtagtaag 1057
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3666 AAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3725
OY 1058 gtaataaagcaaaatgcgaagaaattcaagaaga 1094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3726 AAGATTAACCCCAAAACAAAGATTAAGGTATAGAACAGAAAAGAAAAGTAAAGTAAAG 3762

```

```

RESULT 10
AC024253 179553 bp DNA RTG 07-JUL-2000
LOCUS Homo sapiens chromosome 6 clone RP11-758C19, WORKING DRAFT
DEFINITION AC024253.3 gi:8568958
VERSION AC024253.3 gi:8568958
KEYWORDS HTGS, PHASEL, HTGS, DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 179553)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179553)
AUTHORS Waterston, R.H.
TITLE Direct Submission

```

JOURNAL COMMENT

Submitted (28-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 On Jun 16, 2000 this sequence version replaced g1.7235339.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 ----- Project Information -----
 Center project name: H.NH0758C19
 ----- Summary Statistics -----

Sequencing vector: M13; 100%
 Sequencing vector: plasmid; 0%
 Chemistry: Dye-primer ET; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 175235 bases at least Q40
 Consensus quality: 176522 bases at least Q30
 Consensus quality: 177171 bases at least Q20
 Insert size: 198000; agarose-fp
 Insert size: 178453; sum-of-contigs
 Quality coverage: 4477.53 in Q20 bases; agarose-fp
 Quality coverage: 4.99 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 2275 2274: contig of 2274 bp in length
* 2275 2274: gap of unknown length
* 2375 4601: contig of 2227 bp in length
* 4602 4701: gap of unknown length
* 4702 8520: contig of 3819 bp in length
* 8521 8621: gap of unknown length
* 8621 16434: contig of 7814 bp in length
* 16435 16534: gap of unknown length
* 16535 24088: contig of 7554 bp in length
* 24089 24188: gap of unknown length
* 24189 32883: contig of 8695 bp in length
* 32884 32983: gap of unknown length
* 32984 43009: contig of 10026 bp in length
* 43010 43109: gap of unknown length
* 43110 52414: contig of 9305 bp in length
* 52415 52514: gap of unknown length
* 52515 78544: contig of 26030 bp in length
* 78545 78644: gap of unknown length
* 78645 107099: contig of 28455 bp in length
* 107100 107199: gap of unknown length
* 107200 133817: contig of 26518 bp in length
* 133818 179553: gap of unknown length
* 179553 45736: contig of 45736 bp in length.

```

FEATURES

```

SOURCE 1..179553
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="6"
        /clone="RP11-758C19"
        1..2274
            /note="assembly_name:Contig6"
            2375..4601
                /note="assembly_name:Contig7"
                4702..8520
                    /note="assembly_name:Contig8"
                    clone_end:77
                    vector_side:right"
                    misc_feature 8621..16434
                        /note="assembly_name:Contig9"
                        16535..24088

```


polyA_signal
/product="dntpase"
/protein_id="AAA46131.1"
/db_xref="GI:330963"
/translation="MPYVPELYRFEEDOTETISPARASNLQDINNNILLVNGCVT
IYSTGILIPKERSFAFIIYXGKSKAIPTHGTIGDPGEGELIVLANKTEDTILFEN
DIRAVYLVAVYGVPLRHIDYSDLCPPRYSKNGFDLYLDTQVYAPVBNRNVATCCP
NOLKCYKPIIPGRSLKAGLEIKYKSPNOQLIYFNATQSYTQVAPRRVIAQVNS
HKHHPITLITRKAKPIWHISENITKISWANSYQDITKIPVQDEKLTSSKQTSQMSR
SDAGSGSSGLM"
5761..5766
/note="EDRF3 ORF"
complement(5793..5798)
/note="EDRF1 ORF"
complement(5797..6399)
/note="first methionine codon in the EDFL1 ORF"
/codon_start=1
/protein_id="AAA46132.1"
/db_xref="GI:330984"
/translation="VLCOLHSGNPGRKRRQKYPPLDTSSCPDRMKLEIFIAQPEV
FYGDLLQSDSASLQALILAVQDLNNTIDYLRKDEASVESCYSVOQITLKSV
RIILRTMISLVAVGVITDDALKTLIDRAGESIALIEMKALATAVQDLNSTRAMANE
INNIISAEHKKOILVNSEVPLICEETLAAKRTTPAVSI"
6378..8885
/note="first methionine codon in the EDRF4 ORF"
/codon_start=1
/function="required for DNA replication"
/protein_id="AAA46133.1"
/db_xref="GI:330985"
/translation="MNEATEAMPKEKFLVATDGSABEITDILAGTQNNAFYSVLEN
CIYTEKVIYITGLPAKPKGGGDKCLQVQALIDELALPELFYKPLKANDILHY
IDFKAARNKPIDIDISTNPKSPKTHSGDLKSATWFRATFVNSLKLKATISSPIM
MTTGESVNPPLLAITFEPEQACNTINIIFHLSISEKLGSLAITYITFELGCV
CSTSDITAPALFVYCHIKLARDLSQALDITDILRQGLMSODLVHYLYLSPK
QCLMNDITKTSHLINSDNIHFVEVEVLAQSDLENRKMILYYNRSYVLYKTYIR
CIHPLDILGVAPDOCTSFYVWAGOSKNVHNLINVTNHPINISSEDLNGLDIAAD
SEENNDIKCQVENSCKVPYTRGELKTYFYVYONDILKNTSYVILAPHOENRTM
LKDITITSISKETIFISMLTIRDQKLSREHYNPRLPVFNLDVLDLHISEHT
DEIVYLCCLSLILELQALGPVDIDITHYFPERKTCERPMNIDKREKFCYTRK
LGFRITPLPGVLLGSPNPISEVNLINRTIKIDKILAMPLIMETDGPFDGYTH
KGRVAYRIHTKYVNSGSLREKLKLEFHPVDNKLQYVMSDPINMLYSHNPEV
KOLKAYOIAIDNENEILQKAOALPQNNHVAVERISASMSITDVAEFAWPLREE
LIKILSEHYSQPEHVSFASNGTITIKTILSGNPSCLNTRKHLKQSVRIETSLHL
TPDNCVITITLHSCFASNCNSKCIAMHSVAPITDK"
8881..8886
/note="EDRF4 ORF"
complement(9074..9079)
/note="EDRF5 ORF"
complement(9082..10320)
/note="5' end of coding region for EDRF1 undetermined."
/codon_start=1
/product="transactivator"
/protein_id="AAA46125.1"
/db_xref="GI:330986"
/translation="YGLISDDPDPSNSSSDEFECSQPIKHSQVYASPPSPPEP
SPDVASSTNLRKRNQRPITWEHOSPLSRTKSPRRKRRISNSTSRKTS
WADRVKEAAQORSPRRKPYSHPRNQLPNGPPAPPLKELDILSPKSPKLF
LYVPSLPQEKETNDKLYVLAAGRAMADVYISKQALANLPPKPLSLIYAVYN
PPVYSPKERTIASNLCTVAAFGDESCMOXIKLDEITSENGQIIDEYSALLS
OLVHKIKLPCROCKMORTPODHIYKQVCTYLSNNNLILAVRIQLOSVIKSPVILL
AVANOLAAITCKRNETOLYSHCKRLNEHYRQVKNLHNSLVQHNKRPSSCATY
TRATYGAATITGTEFFLEFQ"
10425..10430
/note="EURF1 ORF"
complement(10679..11752)
/note="first methionine codon in the EURF5 ORF"
/codon_start=1
/protein_id="AAA46134.1"
/db_xref="GI:330987"
/translation="MQCEMTQWVLNIRASQALATPEIWCIFKSLTFSTLSL
QISIVYANVLAHVTIICCYITETKESKESITAGALGALPMLPTGISLWMAPIY
ILPEFYVNCGLSLWPIIDIVYIGPFIYKRYKRGALNAMLVYVYANVIRVYISV
IEMCFPYVILCKGCVALKNPFHEHPYENITLITCAKRYTKTQTEVKNKQIVYAN
LGAIVLLAVASVSVQRLDIFVQNTLFLFYSMLCCCKESLPSSNALCVLEET

Query Match	Best Local Similarity	Score	DB	Length
Matches 211; Conservative 0; Mismatches 245; Indels 0; Gaps 0;	46.3%;	Pred. No. 0.14;		
polyA-signal	complement(11147, .11155)			
CDS	/note="orf" ORF"			
	complement(11749, .12855)			
	/note="first methionine codon in the EELF4 ORF"			
	/codon_start=1			
	/function="required for DNA replication"			
	/product="DNA-binding protein"			
	/protein_id="AA046135.1"			
	/db_xref="GI:330988"			
	/translation="MADPOTHPSEVCGFEPVCOEHTKRYNHKARVNDGIIIDEGITSSGSLRNSTSGEGITLSFKTNVNVSEYFPCNHPPEVSVSFRTSGGNTLTRELFEGCVKATLTFESNRSGKPEETIKALEYCDKVSSTHTSTPGILPEYENVSTSVLSRVLTSPKTLAMLOKMLREBKARKTYSVMNSTLAVLLHVEDSTVDMIVDMSLCNNRSKMGVQNDVOTKVNITLALSALGICKIPGFVPCIKFESEVEVCGTPVKLGSMVDTSLQVLEFCTPFLNADPTCAAVSDPISPSACLTADPADTOEPTLSNCODSSVNSPALPKRRERDINKSKRIATNPPL"			
polyA-signal	complement(11774, .11779)			
	/note="orf" ORF"			
polyA-signal	complement(11174, .11779)			
	/note="orf" ORF"			
CDS	complement(12966, .13883)			
	/standard_name="ribonucleotide reductase beta subunit"			
	/db_xref="1.17,4.1"			
Query Match	1.7%;	Score 64;	DB 73;	Length 43656;
Best Local Similarity	46.3%;	Pred. No. 0.14;		
Matches 211; Conservative 0; Mismatches 245; Indels 0; Gaps 0;				
polyA-signal	complement(11147, .11155)			
CDS	/note="orf" ORF"			
	complement(11749, .12855)			
	/note="first methionine codon in the EELF4 ORF"			
	/codon_start=1			
	/function="required for DNA replication"			
	/product="DNA-binding protein"			
	/protein_id="AA046135.1"			
	/db_xref="GI:330988"			
	/translation="MADPOTHPSEVCGFEPVCOEHTKRYNHKARVNDGIIIDEGITSSGSLRNSTSGEGITLSFKTNVNVSEYFPCNHPPEVSVSFRTSGGNTLTRELFEGCVKATLTFESNRSGKPEETIKALEYCDKVSSTHTSTPGILPEYENVSTSVLSRVLTSPKTLAMLOKMLREBKARKTYSVMNSTLAVLLHVEDSTVDMIVDMSLCNNRSKMGVQNDVOTKVNITLALSALGICKIPGFVPCIKFESEVEVCGTPVKLGSMVDTSLQVLEFCTPFLNADPTCAAVSDPISPSACLTADPADTOEPTLSNCODSSVNSPALPKRRERDINKSKRIATNPPL"			
polyA-signal	complement(11774, .11779)			
	/note="orf" ORF"			
polyA-signal	complement(11174, .11779)			
	/note="orf" ORF"			
CDS	complement(12966, .13883)			
	/standard_name="ribonucleotide reductase beta subunit"			
	/db_xref="1.17,4.1"			
Query Match	1.7%;	Score 64;	DB 73;	Length 43656;
Best Local Similarity	46.3%;	Pred. No. 0.14;		
Matches 211; Conservative 0; Mismatches 245; Indels 0; Gaps 0;				
polyA-signal	complement(11147, .11155)			
CDS	/note="orf" ORF"			
	complement(11749, .12855)			
	/note="first methionine codon in the EELF4 ORF"			
	/codon_start=1			
	/function="required for DNA replication"			
	/product="DNA-binding protein"			
	/protein_id="AA046135.1"			
	/db_xref="GI:330988"			
	/translation="MADPOTHPSEVCGFEPVCOEHTKRYNHKARVNDGIIIDEGITSSGSLRNSTSGEGITLSFKTNVNVSEYFPCNHPPEVSVSFRTSGGNTLTRELFEGCVKATLTFESNRSGKPEETIKALEYCDKVSSTHTSTPGILPEYENVSTSVLSRVLTSPKTLAMLOKMLREBKARKTYSVMNSTLAVLLHVEDSTVDMIVDMSLCNNRSKMGVQNDVOTKVNITLALSALGICKIPGFVPCIKFESEVEVCGTPVKLGSMVDTSLQVLEFCTPFLNADPTCAAVSDPISPSACLTADPADTOEPTLSNCODSSVNSPALPKRRERDINKSKRIATNPPL"			
polyA-signal	complement(11774, .11779)			
	/note="orf" ORF"			
polyA-signal	complement(11174, .11779)			
	/note="orf" ORF"			
CDS	complement(12966, .13883)			
	/standard_name="ribonucleotide reductase beta subunit"			
	/db_xref="1.17,4.1"			
Query Match	1.7%;	Score 64;	DB 73;	Length 43656;
Best Local Similarity	46.3%;	Pred. No. 0.14;		
Matches 211; Conservative 0; Mismatches 245; Indels 0; Gaps 0;				
polyA-signal	complement(11147, .11155)			
CDS	/note="orf" ORF"			
	complement(11749, .12855)			
	/note="first methionine codon in the EELF4 ORF"			
	/codon_start=1			
	/function="required for DNA replication"			
	/product="DNA-binding protein"			
	/protein_id="AA046135.1"			
	/db_xref="GI:330988"			
	/translation="MADPOTHPSEVCGFEPVCOEHTKRYNHKARVNDGIIIDEGITSSGSLRNSTSGEGITLSFKTNVNVSEYFPCNHPPEVSVSFRTSGGNTLTRELFEGCVKATLTFESNRSGKPEETIKALEYCDKVSSTHTSTPGILPEYENVSTSVLSRVLTSPKTLAMLOKMLREBKARKTYSVMNSTLAVLLHVEDSTVDMIVDMSLCNNRSKMGVQNDVOTKVNITLALSALGICKIPGFVPCIKFESEVEVCGTPVKLGSMVDTSLQVLEFCTPFLNADPTCAAVSDPISPSACLTADPADTOEPTLSNCODSSVNSPALPKRRERDINKSKRIATNPPL"			
polyA-signal	complement(11774, .11779)			
	/note="orf" ORF"			
polyA-signal	complement(11174, .11779)			
	/note="orf" ORF"			
CDS	complement(12966, .13883)			
	/standard_name="ribonucleotide reductase beta subunit"			
	/db_xref="1.17,4.1"			
Query Match	1.7%;	Score 64;	DB 73;	Length 43656;
Best Local Similarity	46.3%;	Pred. No. 0.14;		
Matches 211; Conservative 0; Mismatches 245; Indels 0; Gaps 0;				

THIS PAGE BLANK (USPTO)